



# DATA ANALYSIS & VISUALISATIONS OF IPL

# MOTIVATION

The art and practice of visualizing data is becoming ever more important in bridging the human-computer gap to mediate analytical insight in a meaningful way.

—*Edd Dumbill*

# ARCHITECTURE



# TECH STACK

LOADBALNCER  
NGINX S3 COMPUTE  
AWS REDIS CLUSTER  
EC2 STORAGE  
PYTHON SERVERLESS  
ELB LAMBDA  
KIBANA DNS  
DYNAMODB  
ELASTICSEARCH  
GITHUB TRIGGER

ETL

# EXTRACT

- Data has been sourced from multiple areas
  - Scrapping popular cricketing websites
  - Scrapping wiki
  - Google API – Geo points
  - Kaggle competitions
- All the data is further stored into S3.

# Cont...

- Raw data received in S3 is pushed into AWS DynamoDB.
- S3 event invokes AWS Lambda which does the data parsing before it is rested in DynamoDB.
- Three tables – Player, Matches, Deliveries
- Data in DynamoDB acts a source of truth for all the further operations.

# TRANSFORM

- All the semi parsed data is transformed into a meaningful entity - JSON.
- Triggers on DynamoDb would invoke AWS Lambda whenever a new entry is added into DynamoDb.
- AWS Lambda transforms the data into a meaningful patterns, which are further loaded into elasticsearch cluster.
- AWS Lambda would fetch additional geo data through Google API
- AWS Lambda uses Redis for a quick key-value mapping lookup.



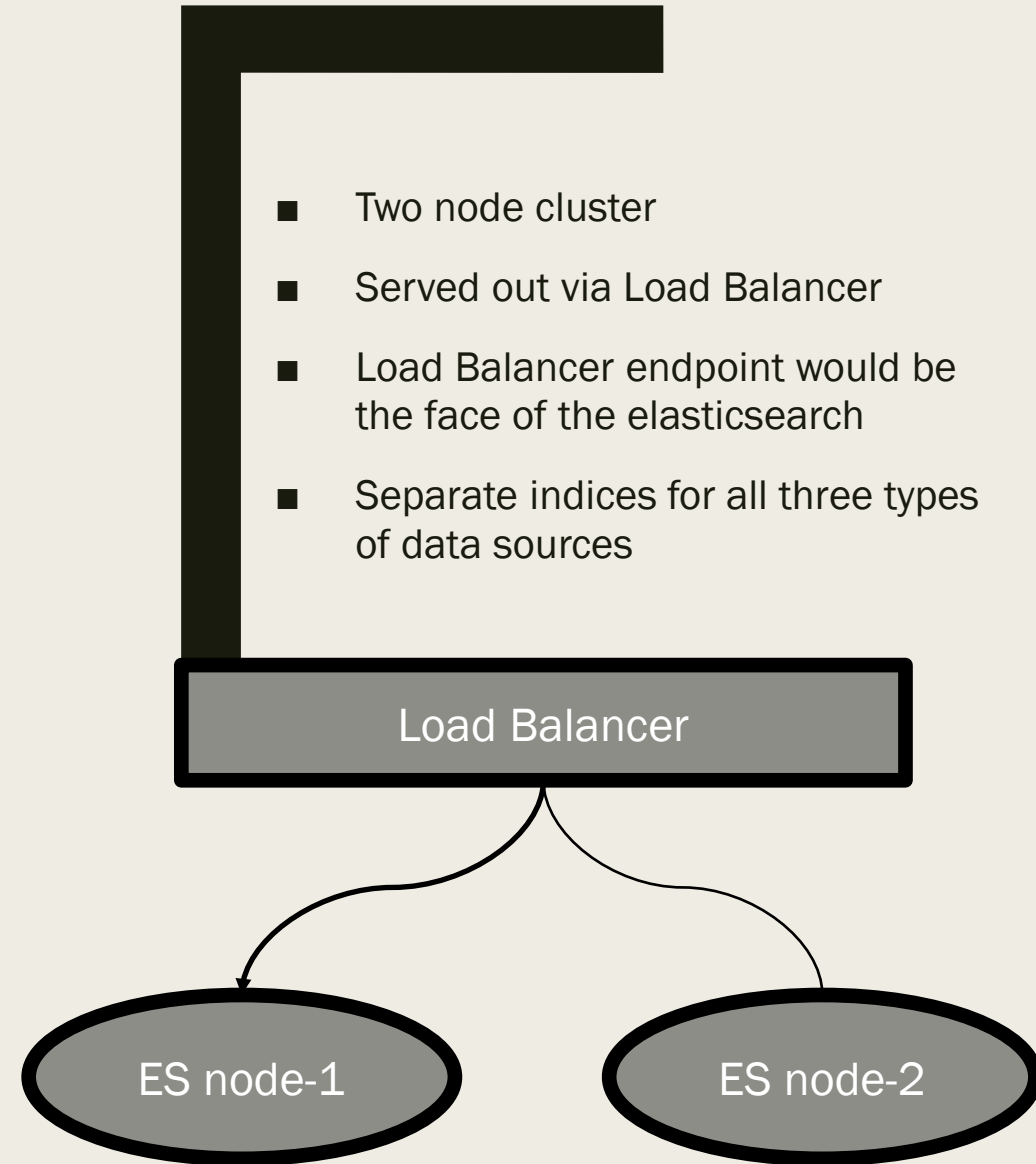
# LOAD

- Elasticsearch indexes all the incoming data from lambdas.
- Data on elasticsearch is split on the nodes in the cluster.
- All the three formats of the data are stored in different indices.
  - i. matches
  - ii. players
  - iii. deliveries



# elasticsearch

- Two node cluster
- Served out via Load Balancer
- Load Balancer endpoint would be the face of the elasticsearch
- Separate indices for all three types of data sources

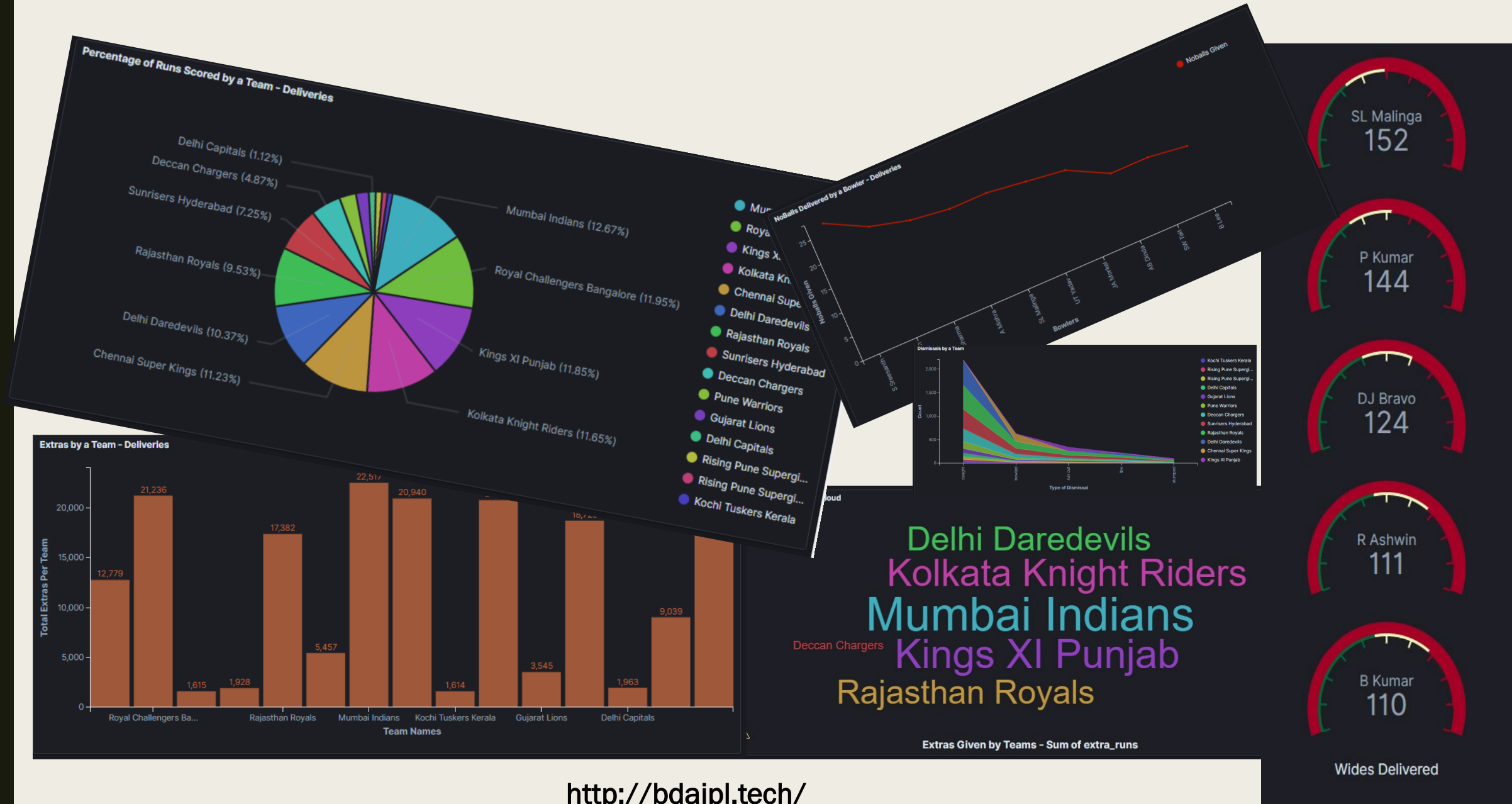




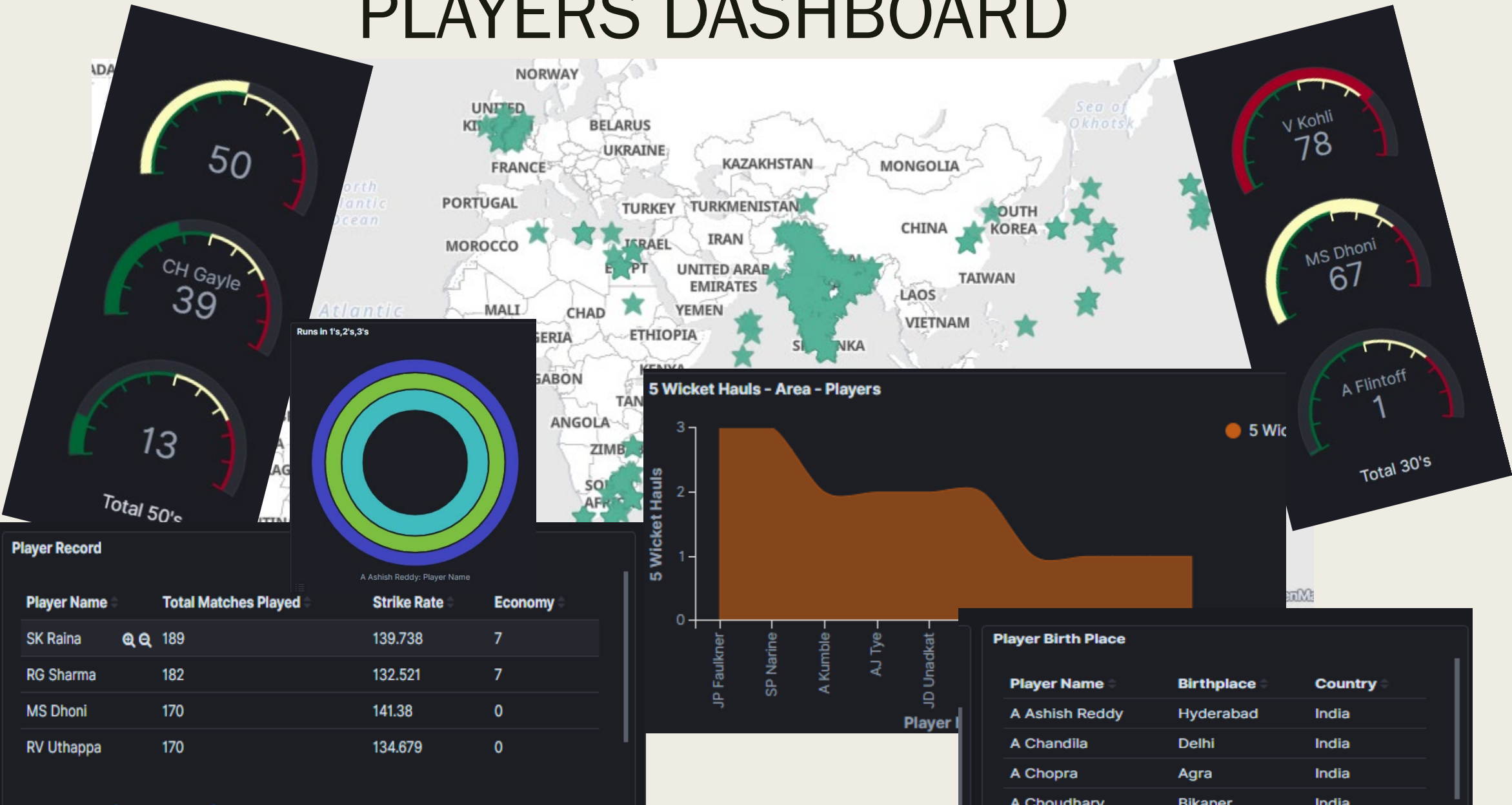
# kibana

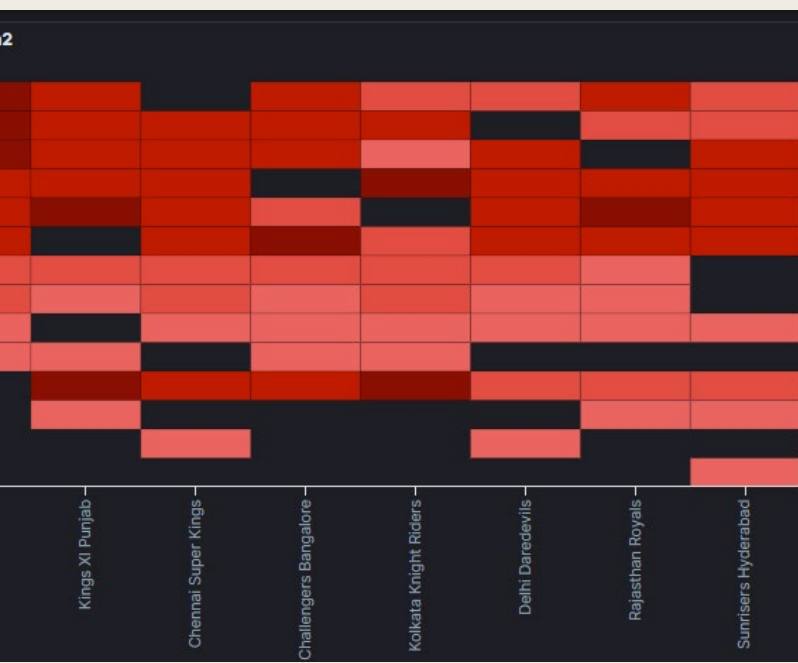
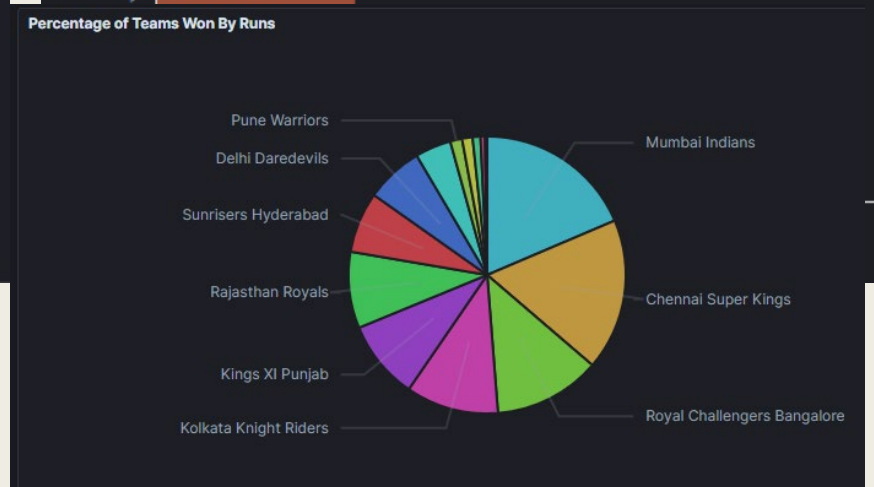
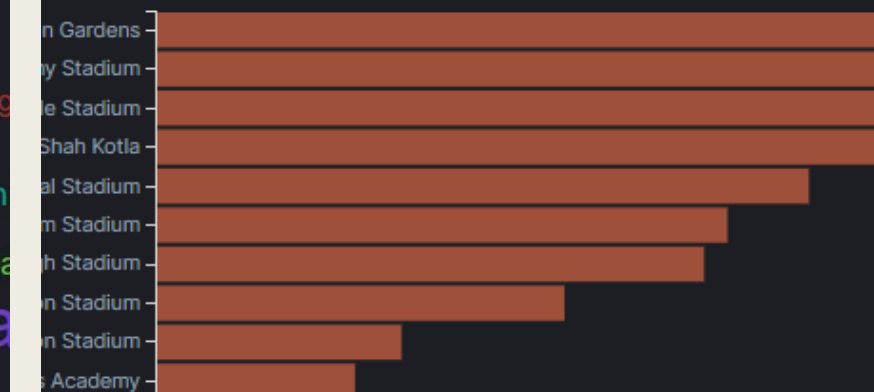
- Face of the project
- Dashboards are generated separately for each index.
- Visualizations such as Geo Tagging, Heatmaps, Custom Metrics, Pie Charts, Bar Graphs
- Filters applied on a visualization is applied across the dashboard

# DELIVERIES DASHBOARD



# PLAYERS DASHBOARD





# MATCHES DASHBOARD



- Port Forwarding
- Forwards request received on 80 to the Kibana's listening port
- Making Kibana as face of the application





THANKS